DOCUMENT RESUME

ED 206 659 TM 810 551

AUTHOR Ellett, Frederick S., Jr.
TITLE Criterion-Referenced Measurement: Redirections.
PUB DATE 15 Apr 81
NOTE 60p.: Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Academic Ability; Behavior Theories; *Criterion Referenced Tests; Induction; Probability; *Research Methodology; *Test Theory; *Test Validity; True Scores
IDENTIFIERS Empiricism

ABSTRACT

Basic issues in criterion-referenced measurement are addressed. In section II, issues involved in determining what a person does and can do are considered. A preliminary analysis of "can" is given which shows that there are several important senses of "can". In section III, results of an analysis of "ability" are given. The analysis is used to criticize current forms of "inductive" inference which employ the concept of "true score". It is argued that using the concept of "true score" lacks justification, and the difference between mathematical probability and inductive probability is explained: that is inductive probability does not conform to the normal multiplication and conjunction principles. Good judgment is shown as being centrally involved with all inductive inference. Alternative Bayesian and non-Bayesian approaches which place various, additional conditions on one's judgments are explored. In section IV, a description of the almost universally employed and traditional logical empiricist method is presented. It is stated that criterion referenced measurement (CRM) theory ought to employ a different method: imaginative participation. Throughout the essay implications are drawn for future CRM theory development and for the training of teachers. (Author/GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED206659

NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Criterion-Referenced Measurement: Redirections

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
F. S. Ellett, Jr.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Frederick S. Ellett, Jr.
University of California at Los Angeles
Graduate School of Education
405 Hilgard Avenue
Los Angeles, CA 90024

April 15, 1981

TM 810 551

Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, California, April 1981.

Criterion-Referenced Measurement: Redirections

Frederick S. Ellett, Jr.
UCLA

ABSTRACT

This essay addresses basic issues in criterion-referenced measurement. In section II, (1) issues involved in determining what a person does and can do are considered. It is also argued that the question about using content or construct validation itself must be rejected, for there is no basis for drawing the distinction between content and construct validation. (2) A preliminary analysis of "can" is given which shows that there are several important senses of "can."

In section III, (1) results of an analysis of "ability" are given. (2) The analysis is used to criticise current forms of "inductive" inference which employ the concept of "true score" (or "universe score"). It is argued that using the concept of "true score" lacks justification. (3) The difference between mathematical probability and inductive probability is explained. It is shown that inductive probability does not conform to the normal multiplication and conjunction principles. (4) It is argued that good judgment is centrally involved with all inductive inference. (5) It is argued that a Subjectivisitc Bayesian theory of probability radically misrepresents both our intuitions and scientific judgments about probability statements. Alternative Bayesian and non-Bayesian approaches which place various, additional conditions on one's judgments are explored.

In section IV, (1) a description of the almost universally employed and traditional logical empiricist method is presented. (2) It is argued that CRM theory ought to employ a different method--imaginative participation.

Throughout the essay implications are drawn for future CRM theory development and for the training of teachers.

Table of Contents

Page

I. Introduction..... 1

II. What a person does and can do:

The Content and Construct Validity Distinction..... 3

A. Introduction..... 3

B. Content Validity as Domain Definitions:

A Case of Excessive Abstraction.....5

C. Commitment to a Form of Behaviorism..... 6

D. Commitment to the Logical Empiricist

Distinction Between Observational and

Theoretical Statements..... 9

E. Conclusions on the Content-Construct Distinction...10

F. More on "what a person can do": Senses

of "can".....11

III. Remarks on Induction and Probability.....15

A. Introduction.....15

B. Comments on the Analysis of "Ability".....15

C. The Current Forms of "Induction" and "True Score"..18

D. Mathematical Probability vs Inductive Probability..29

E. The Nature and Role of Inductive Judgment.....31

F. Probability and the Subjectivistic Bayesian

Theory.....36

G. Conclusion 42

IV. Redirection of Basic Methodology.....42

A. The Traditional Logical Empiricist Method.....42

B. The Defense of Imaginative Participation.....44

V. Conclusion.....51

BIBLIOGRAPHY.....52

Criterion-Referenced Measurement: Redirections

Frederick S. Ellett, Jr.

UCLA

I. Introduction

The theory of criterion-referenced measurement contains at its core a set of questions and tasks which are profoundly important for education. Recent literature on criterion-referenced measurement provides some very important insights and some very provocative questions for all educators.

One central kind of question that criterion-referenced measurement theory tries to answer has this form: Can a certain person do something? For example, does a certain person understand something? Does he know how to perform a certain task? Has a certain person acquired knowledge of such-and-such? What attitudes does the person hold towards, say, physics? As one well-known CRM advocate has put it:

> Unlike their norm-referenced predecessors, which rely chiefly on the relative status of an examinee's performance in relationship to the performance of those in a (statistical) normative group, criterion-referenced tests are supposed to tell us what it is that examinees <u>can or can't do</u>. (Popham, 1980, p. 15, my emphasis.)

And although,

> we'll be focusing on skills to be measured--one might as readily create tests to assess <u>attitudes</u> or <u>knowledge</u>. (Popham & Lindheim, p. 3, my emphasis.)[1]

The other central kind of question that criterion-referenced measurement theory tries to answer is this: By what method (or methods) can one answer those questions about what a person can (or can't) do, and how are these methods justified?

Anyone with the slightest familiarity with education will quickly recognize that such questions and tasks are profoundly central to justi-

fiable educational theory and practice. Anyone seriously interested in educational theory and practice ought surely to be an advocate of a defensible CRM Theory. Equally, anyone seriously interested in educational theory and practice ought to consider critically any applications of the presently available CRM "theories."

In this essay I address some of the basic issues as they are discussed in current CRM literature. In section II,(1) I consider some of the issues involved in determining what a person does and can do. With respect to the question whether content or construct validation is required, I argue that the question itself must be rejected, for there is no basis for drawing the distinction between content and construct validation. (2) I present some results of a conceptual analysis of "can," showing that there are several important senses of "can" which must be distinguished.

In section III, (1) I present some preliminary results of a conceptual analysis of "ability." (2) I use the analysis as a basis for a criticism of current forms of "inductive" inference which employ the concept of "true score" (or "universe score"). I argue that using the concept of "true score" lacks justification and I suggest we consider the concept of "ability." (3) I explain the difference between mathematical proabability and inductive probability, and I draw out the profound significance of the difference. (4) I provide some elementary remarks about the nature and role of induction in our lives and in science and argue that good judgment is centrally involved with all inductive inference. Contrary to common belief, not only is standard setting centrally involved with judgment, but so is induction. (5) I consider the tenability of a Subjectivistic Bayesian Theory of Probability. I argue that such a view radically

misrepresents our intuitions and scientific judgments about probability statements. I explain and explore the plausibility of some alternative Bayesian approaches which place various, additional conditions on one's judgments.

In section IV, (1) I present a description of the traditional logical empiricist method, the method which is almost universally employed today in CRM theory. (2) I argue that CRM theory ought to employ a different method--what I call imaginative participation.

Throughout the essay I draw implications for future CRM theory development and for the training of teachers.

II. What a person does and can do: The Content and Construct Validity Distinction

A. Introduction

A CRM theory tries to specify the methods by which one can determine whether a person can do something. For example, perhaps one wants to determine whether a person can read printed words aloud; perhaps one wants to determine whether a person knows how to solve quadratic equations; perhaps one wants to determine whether a person knows quantum physics; or perhaps one wants to determine whether a person can think critically.

How can we justify or "validate" these claims? Apparently it is believed that there are combinations of only two possibilities: content validation and construct validation. Here are the brief descriptions from the American Psychological Association, 1966, pp. 12-13 (as reported by Cronbach, 1971):

> Content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn.
> Construct validity is evaluated by investigating what psychological qualities a test measures; i.e., by determining the degree to which certain explanatory concepts or constructs account for performance on the test.

Cronbach added later on:

> Whenever one classifies situations, persons, or responses he uses constructs. The term concepts might be used rather than constructs but the latter term emphasizes that the categories are deliberate creations chosen to organize experience into general law-like statements. (p. 462)

In recent review of the literature, Robert Linn has written that content validity is a necessary condition for criterion-referenced measures; however, it is rarely sufficient (1978, p. 3). Cronbach has written that judgments about content validity should be restricted to the operational, externally observable side of testing. Judgments about the subject's internal processes state hypotheses, and these require empirical construct validation (1971, p. 452).

I want to argue that such statements are not false but completely wrongheaded. The questions only arise because of (1) an extreme abstraction of the "content domain," (2) an unwarranted commitment to a form of behaviorism, or (3) an unwarranted commitment to the logical empiricist distinction between observational statements and theoretical statements. Furthermore, Cronbach's own examples betray his theory. Thus, I will argue that no reason has been given for drawing and employing the "distinction" between content and construct validity.

B. Content Validity as Domain Definition:: A Case of Excessive Abstraction

To add to our APA characterization of content validation, consider this:

> The universe specification defines a class of procedures. In most instances the test developed exhibits or describes one procedure and points to a broader universe that, in his opinion, it represents. An achievement test is said to represent a body of content outlined in the test manual. That outline constitutes the universe definition. To ask, "Are the tasks used in collecting data truly representative of the specified universe?" is to examine content validity. (Cronbach, p. 451)

What is being said here? Is there a way to get beyond such abstractions and jargon? Indeed there is. Suppose you want to determine whether a certain person knows quantum physics? "Domain validity" is concerned both with explicating, delineating, and articulating the conditions under which it would be appropriate to attribute knowledge to some person. (Notice that knowledge counts as a construct for Cronbach, et. al.)

But let's put these into still more concrete forms. The former: what concepts, perceptions, attitudes, beliefs, cognitive skills and expectations are characteristic of a person--call him/her an exemplar--who has that knowledge of quantum physics? I take it this is the question to which philosophers of science, and science curriculum specialists have devoted a lot of time and energy. The latter: under what conditions is it reasonable to "infer" that a person has that knowledge and what is the nature of the "inference?"

It is quite standard to consider in a very abstract way bodies of knowledge, e.g., physical science, as if scientific propositions were simply given like natural objects, and as if the only proper questions for consideration were the logical ones of the relations among scientific propositions and between scientific propositions and, say, evaluative

ones (Hamlyn, 1973). According to Max Black, it is time to remind ourselves of how much is ignored in this approach--it is time to consider science as a system of human activities (Black, 1975b). Thinking of science--or any form of experience--as a system of human activities will require that one grasp the scientist's concepts, perceptions, attitudes, beliefs, cognitive skills and expectations--i.e., those aspects which are characteristic of the scientific point of view. (An adequate philosophy of science--an analysis of what science is--must rest on such aspects. See the comments on scientific thinking below).

Once one replaces questions about abstract universes or content domains by questions about a system or set of human activities, one has a sensible way of answering the questions. Furthermore, one can also see that since human actions, concepts, attitudes, expectation and belief play a crucial role in characterizing such things as knowledge, understanding--and even skills--there is no longer any distinction between content validity and construct validity. For content validity as domain definition is also concerned with "psychological qualities". (Recall that philosophers hold that knowledge entails belief! Contrast with Bloom.)

C. The Commitment to a Form of Behaviorism

One might grant that so far as it goes the previous remarks about conceiving "content", "content domain," or "subject matter" in terms of systems of human activities present an important and helpful rephrasing of some of the tasks. But one might then argue that such systems or sets of human activitues have behavioristic reductions. Once the behavioristic reductions are completed, "construct validity" can be

defined as that which deals with the theoretical or unobservable actions and beliefs, while "content validity" can be defined as that which deals with observable behavior.

Most readers are familiar with Skinner's views, so here I explicate and critize Cronbach's position. (See also Hively, et. al, 1973.) In his well known article "Test Validation" (1971), Cronbach, in his discussion of content validity, has written that:

> Statements on response specification in this chapter follow traditional psychological behaviorism, Kendler (1965), for example, equated "behavior" with publicly observable acts and movements ranging from responses that are defined in terms of their achievements to specific muscular contractions and glandular secretions (p. 1). Kendler clearly was willing to make inferences about internal process but insisted on sharp distinction between observations and inferences. (p. 454) (my emphasis)

Unfortunately, Cronbach leaves a good deal unclear about what "traditional psychological behaviorism" is. If "acts" is not equivalent to "physical movement," then "traditional behaviorism" seems to be merely a careless extension of common sense. Common sense wouldn't say that when a person drools or sweats he is doing something. If "acts" is equivalent to "physical movement," then the statement is somewhat redundant, but at least it states a well-known form of behaviorism--the physicalism of R. Carnap (See Malcolm, 1964). It should be noted that Skinner's view of behaviorism does not allow appeals to physiology. At any rate, both views of behaviorism have been subjected to more than one damaging critique and refutation (Block, 1981; Chisholm, 1957; Geach, 1956; Kenny, 1963; Malcolm, 1964, 1971; Putnam, 1969; Scheffler, 1973; Scriven, 1956).

Cronbach's position is further complicated by the claim that "reads printed words aloud" is a proper "observable" for traditional behaviorism

(p. 454). The most a behaviorist can say in the "observation" language is that the person emits certain noises, sounds and physical movements. A behaviorist then claims that such noises and movement are logically equivalent to "reads printed words aloud." It is precisely this equivalence which critics of behaviorism argue doesn't exist. If Cronbach persists in talking in common sensical, everyday terms without giving the logical-equivalence formulas, then perhaps one is quite justified in dismissing Cronbach's talk of behaviorism as merely empty. For in the end, Cronbach, too, relies on the common sense view.

And in common sense, in the human case, there seems to be a deep and general relationship between belief and desire. In the explanation of human actions, there is a relation of epistemic interdependence between the knowledge of the agent's thoughts and beliefs, desires and capacities. One cannot know that one of these is part of the explanation of a certain action without knowing the other two (Chisholm, 1957; Geach, 1956; Richards, 1971, p. 33). And, in general, individuation of actions depends on the agent's beliefs and intentions. Suppose that a person's hand goes up in the air. Whether he is raising his hand to ask a question or raising his hand to stretch his arm will logically depend on his beliefs and intentions. (Perhaps it was a muscle spasm?) In the end, human actions--what a person is doing--are individuated by such things as his beliefs, intentions, or volitions (L. Davis, 1979, p. 41).

If Cronbach holds to his behaviorism where beliefs, desires, and intentions are "internal processes", "psychological qualities", or "internal states", then he cannot determine what the person has done. In common sense to say that a person read a printed page aloud is to assume that the person has certain beliefs, intentions and capacities.

Certain mental terms and internal states are assumed by the statement. At the least, reading aloud is a speech act.

Given that one must use "construct validation" for mental terms--for internal or psychological states and processes--it appears that showing that the person reads a printed page aloud involves "construct validity." But once it is shown that the distinction between content and construct validation relies on behaviorism, the rejection of behaviorism undermines the basis for the distinction itself.

D. Commitment to the Logical Empiricist Distinction Between Observational and Theoretical Statements

One might grant that so far as it goes the remarks about conceiving "content" or "subject matter" in terms of systems of human activities and the remarks about the impossibility of behaviorism are well-taken. But one might argue that there is another way to draw the distinction. One can rely on the logical empiricist distinction between observational statements and theoretical statements. If one equates statements involving psychological terms with theoretical statements, then "content validty" will involve only observational statements while "construct validity" will involve only theoretical statements. According to logical positivists (Hempel, Carnap, et. al), the distinction well represents the way scientists carry on their activites (see Cronbach, 1971, pp. 444, 476, 417, 475-476, 482; see also Hively, et. al, 1973). As many of these writers hold, a scientific theory consists of constructs and observables, interconnected by statements (Cronbach, 1971, p. 476).

For the logical empiricist, the distinction between observational statements and theoretical statements (statements about theoretical entities or constructs) rests on the explication of the distinction between observational terms and theoretical terms or constructs.

The basic difficulty with the logical empiricist's position is that no such distinction is drawn or employed by scientists themselves. The logical empiricist has not accurately explicated or articulated the practice of science; he has given up a distorted picture of the way a scientist works. Many writers have criticized the doctrine held by the logical empiricist. As H. Putnam (1975) has argued:

(a) If an "observation term" is one that cannot apply to an unobservable, then there are no observation terms.

(b) Many terms that refer primarily to what Carnap would class as "unobservables" are not theoretical terms; and at least some theoretical terms refer primarily to observables.

(c) Observational reports can and frequently do contain theoretical terms.

(d) A scientific theory properly so called, may refer only to observables. (Darwin's Theory of Evolution, as originally put forward, is one example.) (p. 217)

I conclude therefore that there is no scientific basis for such a distinction. And since it was its scientific basis which recommended the distinction to us, there has been no reason given for drawing and employing the distinction. (For futher discussion of these issues see Achinstein, 1969; Ellett, 1980).

E. Conclusions on the Content-Construct Distinction.

No justifiable basis has been given for drawing the distinction between content and construct validation. All talk and discussion in terms of the distinction is meaningless and adds nothing to resolving the central questions and problems of CRM.

If one takes "content validity" to be concerned with domain specification or articulation, then a completely separate set of issues arises. (I shall return to these issues in section IV.B.). As advocates of criterion-reference measurement have stressed, they are

trying to determine whether a person is doing such-and-such or whether a person can do something. They also want to delimit or articulate those methods by which we come to defend such statements (see below IV.B). Thus, the central questions of criterion-referenced measurement are:

(I) What concepts, perceptions, attitudes, beliefs, cognitive skills and expectations are characteristic or essential of a person's doing such-and-such (or of a person's having the ability to do something). (This is a task of explication, articulation, and clarification.)

(II) Under what conditions in what situations is it reasonable to "infer" that a person is doing such-and-such (or can do something) and what is the nature of the the inference.

I'll turn in the next subsection to address one of the central tasks of CRM--determining what a person can (or cannot) do.

F. More on "What a person can do": Senses of "Can"

As many noted CRM theorists hold, criterion-referenced measures are supposed to tell us what it is that examinees can or cannot do. But what can be done varies with the meaning of "can" (Aune, 1967a, 1967b). Here are some important distinctions:

"Can" of ability. Here are two subsenses: (1) To have a skill--"He can speak five languages" or "he can paint lifelike portraits." (2) To have the requisite mental or physical power--"He can solve difficult problems," "invent remarkable machines" or "He can swim a mile."

For these subsenses of the "can" of ability, there is an essential distinction between the possession of an ability and the exercise of an ability. To show that a person lacks an ability is more complicated than to show that he does not exercise it. A failure to perform a certain action implies a man (or woman) lacks the corresponding ability only if both he wants, wills, intends, or chooses to perform that action and his

failure to perform it occurs in relevantly normal conditions. A person need not do what he can do. To determine whether he can perform a certain skill one needs to determine what he will do if, in relevanty normal conditions, he wants, wills, or intends to do certain things. And in general, from the fact that a person wanted to do something and successfully did it in one situation it does not follow he has the ability to do that something. Hitting the bullseye once does not show one is a marksman. (Remember also that skills and powers have degrees! See Toulmin, 1971, ch. 7.)

"Can" of inclination (or probability). Examples: "I was so angry I could have killed him" and "That car could hardly have made the trip."

Even those who advocate physicalism or behaviorism must draw the distinction between the "can" of ability and the "can" of inclination. That is, they must distinquish between a behavioral capacity and a behavioral disposition. To a first approximation, a disposition can be specified by a set (perhaps infinite) of input-output conditionals;

If i_1 obtains, then O_1 is emitted
If i_2^1 obtains, then O_2^1 is emitted
and so on.....

A corresponding first stab at a specification of a capacity, on the other hand, would involve mentioning internal states in the antecedents of the conditionals:

If S_a and i_a obtain, then O_a is emitted
If S_b^a and i_b^a obtain, then O_b^a is emitted
and so on....

where S_a and S_b are internal states. In human beings, such states would include beliefs and desires and working input and output organs at a minimum (Block, 1981).

"Can" of opportunity. "He could have played chess had he known how" and "Come in here where we can talk." A person can be said to have the opportunities to do many things that one is not presently doing. To have the opportunity to do something requires only that one be in a situation such that if, roughly speaking, one wanted to do it, it would be reasonable to expect that one would be successful in doing it, if one were able to do it.

"Can" of possibility. The "can" of possibility has at least five subsenses. Here are the two which are especially relevant. (1) the "can" of physical possibility--"If such-and-such has to happen, then it cannot fail to happen." (2) The "can" of conditional physical possibility--"In a deterministic system everything that can occur is necessitated by something else."

It should be clear, then, that the "can" of ability has direct relevance criterion-referenced measurement, but that the others also play important roles. It is possible that, at a given time, the person had the ability to do something, even if at that time there was no opportunity for him to exercise this ability. Educators must provide opportunities for the exercise of abilities. A person who is asleep or on drugs can have the ability but it is not possible for him/her to exercise it.

Notice that to say that someone can (in the ability-sense) do something is to attribute a certain kind of power to him/her. Suppose that the person is given a printed page and asked to read the page

aloud. To say that the person reads printed words aloud--that he can read printed words aloud-is to attribute a certain ability (or power) to him. Cronbach is right in noting that the fact a person has the ability to read aloud says nothing about how he does it--whether the reader looks and says or whether he sounds the words out. Similarly, the fact that a person can solve quadratic equations says nothing of whether he uses the formula or whether he uses successive approximation. Perhaps a new "stimulus" is required to determine how the person does it; but perhaps only a new "response specification" is needed--ask him to write down his way of solving. Thus, to attribute an ability or power to someone is not necessarily to specify that certain way he has of exercising the ability. If one wants to determine whether a person can solve-equations-by-using-the-formula, one may well have to ask different questions. Yet, to determine whether a person has in a given instance solved the quadratic equation ultimately rests on knowing what the person's concepts, beliefs, and intentions are in that situation. If a person cheated or found the answer by "luck," then he did not solve the equation.

That one has solved a quadratic equation one time , however, does not imply that one can (in the ability or power sense) solve quadratic equations. A very important question is when and why are we entitled to conclude that the person has the ability from seeing that he has done such-and-such many times. I take up this question in the next sections.

III. Remarks on Induction and Probability

A. Introduction

As we have seen, criterion-referenced measurement aims at determining whether a given person can do such-and-such and it aims at explicating and delimiting the methods by which such determinations are justified. Given the second task, it should not be surprising to find a lot of discussion about probability and inductive judgments. It is surprising that so little discussion is spent on explicating the concept of "ability."

In this section I shall make some elementary remarks about the concept of "ability" and show how the current probability and induction approaches to this issue are inadequate.

B. Comments on the Analysis of "Ability"

Here I want to lay out some of the issues and problems involved in the conceptual analysis of "ability." Once one has a better grasp of the conceptual dimensions of "ability," he will be in position to judge the adequacy of certain probability and induction procedures. My work here draws upon the ideas of B. Aune (1967a, 1967b) and L. Davis (1979).

First, a person has an ability for a certain time duration. In education, the hope is not only that a person can solve equations in the class today, but that the person will maintain the ability until the next class, and perhaps long after graduation. Notice that I am not saying the person will use the ability acquired in the class after he has graduated; I am only saying that often we hope he will retain the ability.

If one is interested in analyzing what it means to say "a person has the ability to do an action of a certain type at time t", a

hypothetical prediction of the form "if the person tries to do A, he will succeed" will not do. For the person might have had the ability in the past and not have exercised it. But the hypothetical prediction would not entitle us to attribute the ability to the person when he is not trying. Clearly something else is needed. As a first approximation to "a person had the ability to do on A at t" let's try,

"At t, if P had tried to do an A, he would have done an A."

Notice that this formulation is a subjunctive conditional.

A formulation using "chosen," "wanted," or "decided" in place of "tried" would not serve as a general analysis, for not every action is preceded by a choice or desire. But not every action is preceded by a successful attempt to do an A. Davis gives an example of a person whose doing an A is generated by his trying to do a B. An apparently poor marksman who consistently hits a spot to the left of the bullseye may find that if he tries to aim for a spot to the right of the bullseye he can hit the bullseye. So it seems he does have the ability to hit the bullseye. (Notice that for simplicity the account has not mentioned the role of beliefs and perceptions. Surely a full account must find a place for them.)

Davis accomodates this difficulty by leaving the act-type which is the object of the trying unspecified. (Recall Cronbach's remarks on "reading aloud" and "how it is one does it.") And since the word "trying" suggests an effort or prepatory step taken prior to the action, Davis recommends that a general analysis should use "willing." Thus

"At t, a person P has the ability to do an A" means

(A1) "There is an act-type B such that a t, if P had willed to do a B, he would have done A (thereby)."

Unfortunately, not all kinds of abilities require invariable success. Perhaps the ability to move one's finger is an ability which requires invariable success. If someone had the opportunity to move his finger, tried, yet failed, we should deny that he had the ability on that occasion. Yet a golfer may have the ability to sink short putts or a marksman to hit the target, but miss on occasion. Such an occasional (how infrequent?) miss does not lead us to think that the person has lost the ability or that he did not have the ability.

It seems, then, that different kinds of abilities have different success rates. Careful analysis of the ability in question is needed to ascertain the <u>appropriate</u> degree of success. It may be that only a <u>qualitative</u> and not a <u>quantitative</u> measure of degree of success is achievable. Explication and articulation of the given abilities and kinds of judgments made by those knowledgeable is relevant here.

Following Davis we get the following formulation:

(A2) There is an act-type B such that at t, there was an <u>appropriately</u> high probability that if P had willed to do a B, he would have done an A (thereby).

Notice that this formulation assumes that the person could have done a B, "could" in both the <u>opportunity</u> and <u>physical possibility</u> sense. For example, it assumes that no one was preventing P nor was P drugged.

Although there still are several complications to be taken care of here, it will be helpful to consider one more problem. Its educational relevance should be immediately clear. Often educators say that a

certain person today (t) has the ability to do something tomorrow (t'), or that a certain person today (t) has the ability to do something (an A) after graduation (t'). And here the interval from t to t' may be long. (Compare--"the person will have the ability at t'' but he does not have it now (t).") This leads to:

> "A person P at time t has the ability to do an A at time t' ($t > t'$)"
>
> means
>
> (A3) "There is an act-type B such that at t, there was an appropriately high probability that if P had willed to do B, he would have done an A (thereby) at time t'."

Thus, a complete analysis must involve the time dimension of abilities.

This brief and elementary review of some of the problems in the conceptual analysis of "ability" puts us in a position to consider some of the probability and induction proposals which have been put forward to show a person has a certain ability (to an A).

Before we go on, one must notice that one of the major shortcomings of the current literature is the lack of well-articulated and explicated examples of the kinds of abilities and the kinds of probability and inductive inferences which are appropriate for these abilities.

C. The Current Forms of Induction and "True Score"

In its general form, the inductive task is to find premises P_1, P_2,..., P_n such that one can conclude (probably) that a person at time t has the ability to do an A at time t' ($t > t'$). Thus, the general form is:

P_1

P_2

.

P_n

(Probably) -- there is an act-type B such that at t, there was an appropriately high probability that if the person P had willed to do B, he would have done an A thereby at time t'.

It is quite likely that P_1, P_2, ..., P_n involve characteristics of the person P. Notice also the different uses of "probability." (Black, 1970a, pp. 57-59.)

It is helpful to compare and contrast two "forms" of inductive inference which have been suggested:

(1) Traditional Proportional Induction is inference from the frequency of some character in a sample to the frequency of occurrence of the same character in the parent population (Hively, 1973, p. 14; Black, 1970a, p. 58):

M_1/N_1 A's selected by a stated procedure S are B.

M_2/N_2 A's are B.

(2a) Proportional Induction is inference from the frequency of some character in a sample to the frequency of occurence of the same character in the infinite population (Hively, 1973; Linn, 1978; Millman, 1972, 1973).

M_1/N_1 is the proportion of items selected by a stated procedure S are correct.

M_2/N_2 is the proportion of items in the infinite-universe the person could answer correctly if they were all answered.

(2b) Proportional Introduction is inference from the frequency of some character in a sample to the frequency of occurrence of some character in the infinite population (Cronbach, 1971; Novick and Lewis, 1974).

M_1/N_1 is the proportion of items selected by stated procedure S which the person correctly answers.

M_2/N_2 is the proportion of items in the infinite-universe (domain) the 2 person would answer correctly if they were all answered.

In cases 2a and 2b the proportion M_2/N_2 has been called the "true functioning level," "true score" or "universe score" by various writers.

I shall take it that the writers intend the term "true score" to relate to what we ordinarily mean by ability (or capacity). If they believe there is no such connection, then the relevance of their work to real educational problems is very much in doubt.

Given my earlier remarks about the concept of "ability," we can quickly reject proportional induction (2a) because it employes the term "could" in its conclusion. Part of the problem in showing that a given person could do an A (say, could solve quadratic equations) is to analyze the proper sense of "could do." But 2a turns out to be circular.

The First Interpretation

So let's consider 2b. The first interpretation holds that the conclusion to 2b is a counterfactual conditional--a conditional which implies that the antecedent is false (or that it is never fulfilled). As I argued earlier in my analysis, statements which attribute an ability to a certain person P are subjunctive conditionals. Although this is a difference, it does not get the heart of the problem. (Remember, I argued that a hypothetical prediction or open conditional formulation of the form "if he tries to do A, he will succeed" was not an adequate analysis of ability--it would not cover statements about a person who might have had the ability in the past and who might not hve exercised it. Thus the subjective conditional was meant to cover both "counter factual" and "open" situations.)

One of the shortcomings of the "true score" or "universe score" interpretation of ability is that there appears to be no way to get the time dimension into the formulation. Afterall, some times in their lives, people lack certain abilities; some times people have certain abilities; and, unfortunately, these same abilities can diminish or be lost as time goes on. To attribute an ability to someone is to suggest that he have the ability for some (unspecified) interval of time. But since we are dealing with such an extreme counterfactual situation, it is very hard to see how the time aspect can be encorporated. Are we to assume that the person would answer a possibly infinite number of items (or questions) in a finite period of time? How can we attribute the ability to him until he has finished all the items? How does it "if the person were to run help to replace "A person now has the ability to run a four minute mile" with "If a person were to run an infinite number of mile-tracks, then he would run a very high proportion of four minute miles"?

Clearly, "A person now has the ability to run a four-minute mile" does not mean and does not entail "if the person were to run an infinite number of mile-tracks, then he would run a very high proportion of four-minute miles." And, "A person now has the ability to move his finger in this situation" does not mean and does not entail "if the person were to (try to) move his finger in an infinite number of situations like this one, he would have a very high proportion of successful movements."

The major problem is the following. Given that such statements suggest widespread and major violation of scientific laws, it is hard to show that such counter-factuals have any truth-value at all. (Lewis, 1979, p.472). Alternatively, given that the antecedent of the conditional

is true only if there are major violations of scientific laws, it is quite unclear whether one is entitled to assert in the consequent clause that anything would happen. Given such an antecedent clause, perhaps one is "entitled" to assert anything in the consequent clause. Surely, this is absurd. We want to be more discriminating in choosing among consequents of the conditional than this. (Berofsky, 1977, p. 107; Foster, 1975, p. 491). In the face of such difficulties it is best to consider a different interpretation.

Furthermore, it is mistaken to try to employ such a premise in an argument similar to the traditional deductive or statistical syllogism of the form:

M/N C's are B (where $M/N > \frac{1}{2}$)

A is a C
A is a B

To make the argument parallel it would have to run:

M/N would be the proportion of items correct if the person were to answer all of the infinite number of them.

This is one of the items answered (at time t')

The person has answered all of the infinite number of them.
This item answered (at time t') is correct.

The third premise is needed to make use of the conditional in the first premise, yet the third premise is obviously false! Thus, although it has a form similar to a "valid" inductive inference, the argument is unsound or unintelligible if interpreted in such a manner. (See also, Levi, 1980, sec. 11.9, 12.12).

The Second Interpretation

But there is a second general way to interpret such inferences. And here the argumentation involves not only finite or very large

populations which is the domain of statistics, but also infinite (countable and uncountable) populations, which is the domain of modern probability theory. Consider again our formulation (A3) of what it is to say an person had the ability to do A:

> (A3) There is an act-type B such that there was an (appropriately high) probability that if a person had willed to do a B, he would have done an A (thereby).

The key to the second general way of interpreting the arguments is the conceptualization of probability. In particular, given that certain conditions hold, it is claimed that the concept of probability is logically related to the relative frequency of certain outcomes. Thus, the task is to relate the proposition,

> "The relative frequency (or proportion) of items selected by a stated procedure S that are correct (i.e., cases of doing A) is M/N.

to the proposition

> "The probability that a person if he had willed to as a B, he would have done an A (thereby) is P.

Notice that the number P need not be a rational number.

It is crucial to notice that when we are considering continuous cases (rather than the discrete or finite) the probabilities are often represented in the newly developed branch of measure theory. For probability, this development culminated in the axiomatic theory of Kolmogoroff, today the recognized foundation of probability theory in standard physics. (van Fraassen, 1980, p.167). Of importance is the

fact that the axiom of finite additivity is replaced by the axiom of countable additivity. Furthermore, it turns out the domain of definition of probability is both a field and a Borel field (A field is a family of subsets of the sample space K which includes K and which is closed under the operations of intersection, union, and complement. A Borel field (also called a sigma-field) is a field which is closed under countably infinite unions.) Hence, probability is construed as a function which maps sets (in its domain) into real numbers. Probability is construed as a measure. (Some writers hold only that probability be defined on Borel sets or Lebeasque measurable sets. Some don't require countable additivity. See I. Levi, 1980, pp. 104, 122; de Finetti, 1972, pp.90-94.)

Certain writers have argued that a probability space should be interpreted as a model of an experiment: the probability attaches to the possible outcome of the experiment (or test setup). Whether the model bears any relation to observable reality cannot become apparent if the experiment or test is only done once. The clue to the relation between a model and the world must lie in what it is meant to say about repetitions of the experiment (or test or chance setup). As van Frassen notes the intuition is this: more probable outcome events should happen more often (van Fraassen, 1980, p. 181).

Given these rather technical but necessary points, we are now in position to consider the relationship between probability and relative frequency.

Perhaps one of the better known views is the strict frequency interpretation defended by Hans Reichenbach (see Black, 1970b, p.128; van Fraassen, 1980, p.181). Reichenbach argued that probability is frequency. The assertion P(E) = r means simply that, in the case of all

experiments of the sort in question, the outcome E occurs in a fraction or proportion r of all cases (van Frassen, p.181). Reichenbach asks us to look at this structure: a series $s = s_1, s_2, \ldots s_k \ldots s_n$ of actual outcomes (a long run). We can define:

> #(E, s, n) = the proportion of outcomes of E which occur among the first n members $s_1, s_2, \ldots s_k \ldots s_n$ of s.
>
> relf (E,s) = limit (as n goes to infinity) of #(E,s,n).

Here, relf(E,s), called the relative frequency of event E in the long run s, is the number which the limit approaches as we look at higher and higher numbers n.

Unfortunately, such a relative frequency function for some long runs does <u>not</u> constitute a probability function. First, relative frequency is not countably additive. Here's van Fraassen's example:

> "Let the actual long run be counted in days: 1 (today), 2 (tomorrow), and so on. Let A(n) be an event that happens only on the nth day. Then the limit of the relative frequency of the occurrence of A(n) in the first n + q days, as q goes to infinity, equals zero. But the union of the events A(n)-that is, A(1)-or-A(2)-or-A(3)-or-....has relative frequency 1. That happens every day"(p.184).

Secondly, it can be shown that the domain of definition is not a Borel field. Furthermore, as deFinetti and Rubin have shown, the domain of relative frequency in the long run is not a <u>field</u> either (van Fraassen, p. 184). Therefore, we cannot say that relative frequencies are probabilities. And except in restricted classes of events, probabilities are not relative frequencies either (p.184).

The failure of this approach which tried to use the relative frequency of <u>actual</u> outcomes lead to interpretations of probability which involve

propensity and virtual sequences. For example, Karl Popper put it roughly this way: when we say that outcome E has a probability of 1/2 in experiment C under present conditions, we mean that E would occur half the time, if the experiment C were repeated sufficiently often. (Compare and contrast to Cronbach's definition of "generalized true score" or "universe score", p.449.) The probability of E is not the relative frequency in the actual course of events (which may end to soon or in which the conditions of the experiment vary with time, etc), but it is relative frequency of the sequence of outcomes which would occur if C were repeated indefinitely under the same conditions. But how is this to be tested?

The answer is that the actual sequence must be regarded as random part of the virtual sequence. The comparison would make no sense if we could not regard the actual sequence as a random selection from the virtual one. Popper must be construed as saying that, if in the actual world the relevant conditions do occur infinitely many times, then the actual frequency equals the virtual frequency.

Let us pause for a moment to notice that if such a conceptualization is applicable to test construction, each item (or question) must be viewed as a random sampling. Yet if we ask someone to hit ten golf balls one after the other, there is little reason to think each of the "hits" is a random sampling. Again, consider the situation where we ask some to shoot a gun at the bullseye. In such situations, it is quite likely that in the end of the sequence the person does better than in the beginning. (Consider what would happen if the sequence contained a hundred trials.) To be applicable it must be empirically established that the experiments or trials are random samplings. (I. Levi, 1980,

pp.393, 396.) (It is also arguable that the conditions of randomness and independence assumptions are not valid for typical examples of scientific inference (Hesse, 1975, p.78).

Unfortunately, such propensity interpretations of probability can not be justified. They fall prey to all the difficulties that beset the strict frequency interpretation. There are cases where we find the domain of the definition may not be a Borel field, and indeed not even a field and even where defined, not countably additive (van Fraassen, p.190).

Given that these attempts to show how probabilty is related to the relative frequency of a certain kind of sequence are unacceptable, does it follow that probability is not linked in any way with the frequencies of outcomes? Of course not. Of particular interest is the rather sophisticated "modal frequency interpretation" which van Fraassen develops in response to the inadequacies of the previous interpretation (See his sect. 4.4). (Briefly, van Fraassen attempts to set restrictions on the domain of the function relf(E,s) so that it is a probability measure which in the limit converges to the proper value of probability, P.) I believe van Fraassen's approach is interesting and that it deserves further critical scrutiny.

Nonetheless, recent arguments by I. Levi cast doubt upon the philosophical adequacy of such an analysis. Objective probabilities (or chances) of <u>single</u> trials are not logically or conceptually or incorrigibly linked to chances on <u>sequences</u> of trials of the same kind. Chance predicates are interpreted by Levi to be <u>primitive placeholding</u> predicates (p.260). The chance predicate serves as a placeholder in stopgap explanations. In fully acceptable explanations, such place-

holding predicates would not be allowed to appear in law-like statements. But what qualifies as a fully acceptable explanation depends on the state of inquiry and the programs for explanation to which the investigators are committed (pp.237-238). Levi has argued that chance predicates are no more expendable in inquiry and deliberation than disposition predicates are (Levi, sec 12.9, 12.10, 15.6).

Objective probabilities or chances play a role in evaluating belief (or credal) probabilities for hypotheses about frequencies of a certain outcome on a single trial. Sometimes one wants to evaluate the belief (or credal) probabilities for a hypothesis about a random sequence of certain outcomes of certain kinds of trials. The important point is that saying the selection (or the sequence) is random is saying that the chance of obtaining one member of the population is equal to the chance of obtaining another on each trial of the sequence (pp. 272, 273, 393). Of course, whether the chance of a certain outcomes remains the same over trials of similarly prepared setups is not guaranteed (p.273). But to claim that the trial is random is to presuppose the intelligibily of the probability of outcomes of a single trial.

Whether such a notion of objective probability (or chance), which regards the outcomes of a certain trial setup to be stochastically determined, can play a role in criterion-referenced measurement of ability has yet to be fully established. (See the next subsection.) Yet such a use does provide a promising alternative to using the current concepts "true score" and "universe score." As I have argued, it seems that the concepts "true score" and "universe score" lack both logical and empirical foundation. Clearly further research and argumentation is required concerning the value of objective probability for criterion-referenced measurement

D. Mathematical Probability Versus Inductive Probability

As I have argued earlier, the concept of ability (or 'disposition' in the wide sense) is connected to the concept of probability. In the previous section I have discussed how one kind of Bayesian might use a conception of probability to analyze ability. It must be remembered that all Bayesians assume that any viable conception of probability must obey the multiplication, finite addition, and complimentational negation principles of the calculus of probability. Any conception which satisfies such principles will be called a conception of mathematical probability.

It should be noted, however, that there are writers who claim that there is a conception of probability which differs significantly from the mathematical conception employed by Bayesians. In many places Black has suggested this view when he claimed that 'probability' is related to "forcings," or "influences." (1970b, pp. 98-99, 135). This view is most clearly articulated and defended by L.J. Cohen (1977). It is Cohen's position I will briefly describe here.

As others have argued, Cohen argued that a dispositional statement--i.e., a statement ascribing a dispositional predicate--is a statement about probability. Given my earlier remarks about ability, it can be seen that Cohen's arguments apply to abilities as well as to dispositions.

But Cohen has argued that it is not mathematical probability that is involved. Rather, a dispositional statement is a statement about what Cohen has called inductive probability. (1977, Ch. 23).

Cohen has argued that mathematical probability is different from inductive probability in several ways. Two of the more important differences are that the former but not the latter is governed by the

standard multiplication principle and that the former but not the latter is governed by the standard complimentational negation principle. (Ch. 17, sec. 64.) In other words, the principles governing mathematical probability differ significantly from those governing inductive probability. Here is the relevant multiplicative principle for the probability of a conjunction for inductive probability:

$$\text{If } P_I\ (S'/R) \gtrdot P_I\ (S/R), \text{ then } P_I((S \text{ and } S')/R) = P_I\ (S/R).$$

Here is the relevant negation principle for inductive probability:

$$\text{If } P_I\ (S/R) > 0 \text{ and } P_I\ (\text{not-}R) = 0, \text{ then } P_I\ (\text{not-}S/R) = 0.$$

Suppose we can correctly describe a man being of a solitary but courteous disposition (or, say, a dog as being of a friendly but a greedy disposition.):

> "If he does have a compound disposition, the whole of the compound will be activated in appropriate circumstances, and the same is true of the dog that is both friendly and greedy. If the constituent dispositions were contrued as more or less independent mathematical probabilities, however, the multiplicative principle for the mathematical probability of a conjunction would force us to the paradoxical conclusion that most dispositions were uncompoundable. The product of the mathematical probabilties that are supposed to be implicit in the constituent dispositions might well fall below the threshold at which a disposition is said to exist. Analysis in terms of inductive probability avoids this paradox because the conjunction principle for inductive probability fits in with our ordinary views about the compounding of dispositions. The inductive probability of a conjunction is no lower than that of its least probable conjunct." (p. 329.)

It must not be thought that inductive probability is of little importance. Cohen has argued that inductive probability is used in stating the standards of proof in both British and American Courts. Also, he has argued that we all constantly employ the concept of inductive

probability in reasoning about what has probably, or certainly, happened in a certain situation. (p. 346.) Furthermore, it reconciles philosophy of science with the ordinary man's conception of proof and knowledge (p. 348). Once the role of inductive probability is seen in an adequate account of inductive logic, one can show that the widespread belief that scientific knowledge is impossible is unjustified. (Cohen, sec. 98.)

In the light of these remarks it is sufficient to conclude that Bayesian inference and mathematical probability see unable to provide a complete account of inductive reasoning. Not only has Cohen presented for certain contexts a feasible alternative to the Bayesian account, but he has argued that his account--which employs inductive probability--solves many of the problems and questions which face any view of inductive logic. I shall pursue in another paper the general but profound question about when and where a Bayesian or an inductive probability account is defensible.

E. The Nature and Role of Inductive Judgment

Given that so many of the writers on CRM employ "true score" or "universe score" and given that I have shown no justification has been given for using that concept, would not reversion to the ordinary concept of ability solve the problems. Unfortunately, the answer is no. The literature contains a view of probability and inductive judgment that is subject to serious limitations.

Almost all writers agree that questions of standards (or criteria) are shot through with judgments (Jaeger 1976; Linn, 1978; Messick, 1975; Millman, 1973). Some go so far as to claim that setting standards (or mastery levels or cut-off scores) is ultimately arbitrary. Questions of standards (and values) are imprecise and ill-formulated; answers to them involve judgment and are arbitrary. By way of contrast this suggests

that probability estimates, inductive inference, and Bayesian statistics are not "shot through with judgments," that these are not arbitrary but precise and well-formulated.

Contrary to widespread belief, probability and induction are riddled with judgment. Contrary to widespread belief, certain forms of Bayesian probability are arbitrary. In this subsection I will argue the former; in the next subsection the latter claim.

As I have argued earlier, many writers have tried to use forms of inductive argument which are similar to the traditional forms of proportional induction and proportional deduction (or statistical syllogism). Of course these are not the only forms of inductive argument. Indeed, there is no general agreement about the basic forms, although many writers regard simple enumeration--from a finite number of cases where A_i is B one argues any A is B--as in some sense the most fundamental (Black, 1970a).

My purpose here is not to give a formal defense of a view of inductive argument. Rather my purpose is to remind people about what they already know about inductive argument and the crucial role it plays in our daily lives and in science.

Perhaps this example will help: you wake up and notice that your watch has stopped--your miniature inductive problem is to decide whether it is time to get up yet. Yes, this is an inductive problem; our lives are filled with inductive problems. An important point is that there are intelligent as well as stupid ways of attacking these problems.

What is involved in "giving" a good "solution" to an inductive problem? I believe the following characterization given by Max Black captures well our experience:

> In a good inductive solution, there are present at least the following: Recall and choice of relevant premises; estimation of the strength with which they support the conclusion under examination; conflation of several supporting strands to yield a judgment of the resultant strength of the conclusion; a decision whether a suitable "detachment point" has been reached, or whether suspension of belief is in order. (Black, 1970c, p. 143)

As Black notes, at point after point in the process, the reasoner's flair, skill, good judgment are crucial. Induction is an art, not a science or a system of mechanical rules.

It is important, too, to recognize the role of background knowledge and previous experience. All sound inductive judgment is made against a background of detailed knowledge that activates the skill and expertise built up by repeated exercises of inductive judgment. It may well be that such knowledge can never be fully formulated (compare to Polanyi's tacit or implicit knowledge).

Issac Levi has advocated the view according to which considerations other than logic, language, and the available evidence play a role in the justification for adopting a prior (credal) probability state. The problem under investigation, the rival potential answers identifed for it, their evaluation with respect to informational values, and the index of caution may also play a role (Levi, 1980, pp.315, 424-430).

Concerning scientific thinking, Wesley Salmon has argued that the Bayesian schema comes closer than the hypothetico-deductive schema to capturing actual scientific practice, for Salmon believes that scientists do make substantial use of plausibility considerations, no matter how adept they are at fully articulating them (Salmon, 1980, p.82). Such plausibility judgements of the hypothesis is prior to, or apart from the results of directly testing that hypothesis (p.80). Salmon has argued that most philosophers of science reject the H-D schema in favor of a

Bayesian schema, but that they differ in regard to the nature of the prior probabilities (p.85). Salmon also argued that the plausibilities are based upon empirical experience with scientific hypothesizing, with dealing with hypotheses of similar types (pp 85-86). Apparently, then, background knowledge and experience play an important role in the scientific justification of hypotheses.

Mary Hesse has advocated a modest Bayesian confirmation theory which is more specific than is usual in the literature. She has argued that there are additonal criteria of adequacy (which are called "clustering" "analogy", and "simplicity"). Given such criteria, she argued that confirmation in science should be seen primarily as an application of inference from case to case rather than as a confirmation of universal laws and theories (1975, pp.91-105). Hesse has argued that these patterns of inference depend on our intuitions and unformalized background knowledge (pp.95, 95-99) (Note that her notion of "simplicity" is syntactical. pp.102-103). (See also, H.I. Brown, 1977; P. Thagard, 1978.)

But how do we determine whether a person is a good inductive judge? Although the question is difficult to answer in detail, the common sense answer is that we notice how often he goes wrong without extenuating circumstances. A good inductive judge must not fail too often (Black, 1970c, p. 143).

But, you will say, how often is "too often"? The answer--it takes a good inductive judge to <u>judge</u> an inductive judge--if only because the arguments in question have to be independently appraised. Stop and think about how you have learned to make such judgments, about how to make such judgments about your friends', students', or child's judging. (Why do you think there must be a mechanical rule which requires no judgment?)

At this point many will be quite edgy. Educators have the tendency to regard all "judgments" as problematic, perhaps as arbitrary. There is much that can be said to clear up such confusions and misunderstandings. Let me only note that we learn to become better inductive judges by scrutinizing our failures and retroactively criticizing the trains of thought that produce them. Inductive argument is self-critical in an important sense (Black, 1970c, p. 143; Levi, 1980).

The importance of these remarks is at least two fold. First, by reminding ourselves about the nature and role of inductive argument in our lives and in science, we can get in position to judge the adequacy of the formulized, mechanical rules which various writers have put forward for educational theory and practice. It cannot be emphasized too strongly that sound inductive argument involves, point after point, good judgment and that that judgment can be learned.

Second, there is a tendency in education to want to give the teacher a set of procedures and mechanical rules so that he could make better inductive inferences about what his students know, understand or can do. Yet if these procedures and mechanical rules distort and misrepresent inductive argument, then the teacher may well end up making worse inductive inferences. Furthermore, it seems plausible to hold that certain teachers, by means of past experience and reflection, have become quite adequate inductive judges. It seems plausible to hold that such teachers can not be improved upon by any mechanical procedure. (Why do educators tend to think teachers can only follow recipes?) And perhaps it is more important to remember that one learns how to be a good inductive judge. Why not devise programs that would encourage teachers to develop and to scrutinize their current inductive powers and strategies?

I will not deny that there are difficult questions for this approach to answer. But I maintain the approach is viable and should receive the attention of some of our best scholars and practitioners. Let me add that I am quite supportive of efforts to formulate certain procedures and to develop mechanical rules. These things have a place to play. But it is crucial to remember that the adequacy of such things must ultimately be justified by inductive argument, by sound inductive judging!

One set of problems which a philosopher (or a sociologist) finds extremely urgent is that of clarifying, analyzing, and if need be, "reconstructing" the concepts, strategies and maxims involved in inductive argumentation. Remember that the "begin-all" of such work are our "ordinary" or "preanalytical" concepts and judgments. A theory of induction is meant to articulate and explain our ordinary and scientific inductive judgments. (Black, 1970b; Cohen, 1977; Hesse, 1975; Levi, 1980; Salmon, 1970; Thagard, 1978).

F. Probability and the Subjectivistic Bayesian Theory

To talk about probability is to talk about something related to but distinct from induction. (Indeed probability judgments presuppose some inductive conclusions.) Here I want to consider certain aspects of the concept of probability; I shall not consider inductive arguments per se.

The importance of this inquiry stems from the influence of the views of Melvin R. Novick on educational theory and practice. Novick and his associates have apparently been able to establish their view as one of the major theories of probability in educational theory. Furthermore, it is usually recommended that such views be used to construct educational policy. I shall call the apparent view of Novick, et al., the subjectivistic Bayesian theory of probability. I shall argue that

there is no reason to think the theory has significance for theory or practice.

The chief virtue of this theory of probability, according to both advocates and critics, is that it provides a method of establishing the prior probabilities of various events (Novick & Lewis, 1974; Linn, 1978). Once one has the prior probabilities, one uses Bayes' inverse probability theorem to derive certain posterior probabilities. Using Bayes' theorem is what makes Novick a Bayesian here; he does not commit the quite serious error of assuming equi-probable priors as many others have done. In general, suppose one has a number of hypotheses H_i that are compatible with the evidence E and that one knows the forward probabilities or likelihoods $P(E/H_i)$. If one knows the prior probabilities and uses Bayes' inverse theorem, one can determine $P(H_i/E)$--the posterior probability--the probability that H_i is true given the evidence E. It should be obvious that knowing $P(H_i/E)$ would be extremely useful in educational theorizing and practice.

Interestingly enough, the standard remark about the approach is that "the choice of the prior distribution is a key and controversial aspect of the approach" (for example, Linn, 1978). Nothing else is said in criticism. At this point, however, the evidence supports Novick. He has a plausible case that any defensible procedure of inductive testing of theories presupposes some prior weighing of the theories. He is correct in suggesting that a theorist should try to specify, clarify and articulate the set of priors he is using in his inductive testing of theory. It is a widely accepted doctrine in philosophy of science that any inductive testing presupposes some apriori (in the sense of antecedent)

weighing of theories--a weighing prior to the checking. (Sometimes these orderings are called "plausibility" or "simplicity" orderings.) Recall Salmon's remark that philosphers of science seem to prefer a Bayesian schema to a hypothetico-deductive H-D schema, but they differ about how the priors are to be determined. (See also, Hesse; Levi; Putnam, 1978, p.75.)

However, it is also widely accepted no one has been able to formalize the intuitive judgments of scientists which in practice constitute the "prior probability metric." Here is an area of inquiry where psychometricians, sociologists, philosophers and scientists could do some very important research! But that's another story.

Thus, it is not a telling criticism of Novick that he recognizes the need for and tries to provide a procedure for obtaining the prior probabilities. Those who fail to recognize the need and to take steps to fulfill it most likely deserve the censure.

Whether Novick's view is tenable depends in part upon the nature of the procedure for weighing or getting the prior probabilities and in part upon the additional restrictions he places on the beliefs. I believe that the subjectivistic view is seriously flawed and should be abandoned.

The subjectivistic view holds that the prior probability of a hypothesis is to be identified with the person's strength of belief that the hypothesis is true. (Novick and Lewis, 1974, pp.144, 146, 155.) No philospher has proposed this view, if only because it would make probabilities fluctuate too much to be worth considering. It is also surprising that Novick seems not to require that the person's beliefs about various hypothesis satisfy the addition and multiplication rules of the mathematical

theory of chances. That is, Novick seems not to require that the person's belief form a probability function. Given these limitations, such a view must be rejected by any Bayesian.

L.J. Savage and B. de Finetti have argued for the following additional conditions: the person's beliefs are to be consistent and coherent. A set of beliefs is coherent if it is impossible for anybody to "make book" against him, by accepting odds that will guarantee a net profit. (Of course, there are different intervalist strategies for determining a coherent set of beliefs. See I. Levi, 1980, sec. 9.8.) If $b_s(H/E)$ is the person's degree of belief in H given E, then for different choices of H and E, the person must rectify his beliefs so that they are consistent and coherent. (See Black, 1970b, pp.130-133). Given that a set of beliefs is consistent and coherent, it follows that those beliefs will obey the multiplication and addition rules of the mathematical theory of chances. This result gives the "personalist" access to the usual mathematical axioms and their consequences.

Other writers, however, have argued that we need to supplement consistency and coherence with a principle of direct inference which stipulates how knowledge of chances or (objective probabilities) determine credal probability judgements about the frequency of outcomes of trials on chance setups. (See Levi, 1980; van Fraassen, 1980). It is important to notice that the concept of objective probability (or chance or statistical probability) is not to be interpreted epistemologically. As Levi puts it, chance statements do not describe the cognitive states of agents. They bear truth values and attribute properites or conditions to objects which they have (or lack) independent of the subjective states of knowing subjects. (Levi, sec. 12.9, sec. 12.10, and sec. 16.6; van Fraassen, ch. 6, sec. 2.2.) (For a discussion of other conditions which various

writers require see Levi, 1980. See especially his discussion of "confirmational conditonalization.")

Several writers have argued that one should give up the traditional Bayesian requirement that there be unique and numerically precise prior probabilities. Such writers would allow indeterminate but interval-valued priors. That is, one needn't require that a person believe that the probability of getting a head with coin a on a toss is p, one could allow the person to believe that the probability of getting a head falls in some interval. Given such developments, I. Levi has argued that the standard intervalist strategies for determining a person's probability-beliefs--e.g., the fair bet to avoid a dutchbook--must be rejected. Levi has offered an alternative to such strategies (1980, sec. 9.8, 9.9.)

Given these rather striking modifications in the doctrines of traditional Bayesian approaches, it is unclear whether such developments should also be called Bayesian![2] At any rate, as I've noted above, several authors have argued that background knowledge and experience play a role in determining the priors. (Black, Hesse, Levi, Salmon, Thagard.) Many have argued that the problem under investigation, the rival potential answers identified for it, their evaluation with respect to informational value, and the index of caution may also play a role.

As we continue to learn about the structure of our knowledge about the world and about the reasoning which is useful in acquiring that knowledge, it may be found that background knowledge experience and intuition play an essential role in scientific inference. Yet it may also be shown that such a view does not undermine the objectivity of science so as to reduce it to mere psychology or sociology (Levi, p.27).

Whether a Bayesian approach (or one of its recent developments) can be defended is, of course, still a controverisal question. Many have argued that coherence may not be a requirement of rational belief. (Black, 1970b; Cohen, 1977. See section III.D., above.) If such a criticism could be sustained, all plausible forms of Bayesian theory would be seriously undermined. Furthermore, Paul R. Thagard has argued that certain contextual factors are required to give an adequate account of scientific reasoning (Thagard, 1978; See also Cohen, 1977). He offered three criteria: consilience, simplicity, and analogy. (Compare and contrast with M. Hesse's concepts.) To say that a theory is consilient is to say more than that it "fits the facts": it is say first that the theory fits the facts, and second that the facts it explains are taken from one than one domain (p.82). And at a given time, the leading scientists agree on the division of facts into domains (pp.80,85). A characterization of analogical inference can be given using the concept of explanation. Suppose A and B are similar in respect to P, Q, and R, and suppose we know that A's having S explains why it has P, Q, and R. Then we may conclude that B has S is promising explanation of why B has P, Q, and R (p.90). Thagard has argued that unlike hypothetico-deductive and Bayesian models of theory evaluation, his "best-explanation" view gives an integrated account of the nature and importance of simplicity, ad-hoc-ness, analogy, and variety of instance (p.91)". Notice also that Levi has criticized many of the doctrines of the traditional Bayesian. Do Thagard's criticism's hold against the kind of position developed by Levi? I'll pursue this question in another paper.

X7FSE/A

G. Conclusion

Recent theories and views about criterion-referenced measurement have been advocated for educational policy and in some cases have been put into practice, but serious problems remain. As a conclusion to this section, I believe the following remarks by Isaac Levi have obvious implications for educators:

> "In the wake of recent episodes at nuclear power plants, not only have policies favoring nuclear energy been discredited in the eyes of the public; but the credibility gap has extended to pure scientists and engineers. I do not seek to join the chorus of critics of scientific approaches to the solution of problems. If the serious problems we face are to be solved, they will be solved only by careful and sophisticated inquiry.
>
> But the demands of such inquiry require that experts admit the limits to what they know. Scientists and technologists should not pretend to a knowlege they do not have because a government or a public demands that they be supplied with answers to which there is insufficient evidence. And the public and government should understand and respect the limits on what they can expect of responsible scientists and engineers. They should refrain from putting unreasonable pressures on investigators to subvert their better judgement.
>
> Above all we should beware of epistemologies which permit us to violate this counsel and indulge over tastes for the familiar, simplicity, explanatory power, naturalness, paradigmatic methods of puzzle solving, and the like without regard to the risks of error both in theory and in practice which over indulgences may be incurring (1980, p.444).

IV. Redirection of Basic Methodology

A. The Traditional (Logical Empiricist) Method

For many years educational research and practice has held that the methodology appropriate for the social sciences - and for education - is the same methodology as that of the physical sciences. Many of those interested in criterion-referenced measurement hold the same belief. In this section I will argue that the belief ought to be abandoned--that there must be room for a different (basic) methodology in eduation. (Or at least, we must bring the methodology of the social sciences in line with a proper account of scientific methodology.)

Let's begin with a few quotations from a quite influential educator. Cronbach, in his 1971 Chapter "Test Validation", quoted with approval this remark by Nagel (p. 483):

> "The crucial point is that the logical canons employed by responsible social scientists in assessing the objective evidence for the imputation of psychological states do not appear to differ essentially (though they may often be applied less rigorously) from the canons employed for analogous purposes by responsible students in other areas of inquiry (Nagel, 1961, p. 484)."

According to Cronbach, the logic of confirmation and disconfirmation is the same for constructs in education and social sciences as in natural sciences (p. 477). And such a rationale has been generally accepted by psychological testers (p. 480). In general, the method is the hypothico-deductive model.

Those in the positivist tradition have long held such a position. J.S. Mill held that the backward state of the Moral Sciences (or the social sciences) could only be remedied by applying to them the methods of Physical Science, duly extended and generalized (1881, BK VI.). Nagel has written that there appears to be no good reason for claiming that the general pattern of explanations in historical inquiry, or the logical structure of the conceptual tools employed in it, differs from those encountered in the generalizing and the natural sciences (1953). Here one can assume his remarks apply to educational inquiry.

Even those who grant that there may be a method which is unique to the social sciences go on to draw the distinction between the logic of discovery and the logic of confirmation. They may then grant that the method belongs to the logic of discovery - if they grant that one exists - but they steadfastedly hold to the claim that any method needs to be checked by the scientific method (SM). The scientific method is identified with the logic of confirmation (Rudner, 1966, p. 73).

These views are typical of those who hold that the only kind of knowledge is scientific knowledge , knowledge acquired by the "scientific method (SM)". Anything that can be known in educational inquiry can be known only by these methods. The hypothetico-deductive method, its simpler inductive relatives, and of course, observation are all methods there are for acquiring non-demonstrative knowledge (Putnam, 1978, p. 67; See also Hesse, Salmon, and Thagard). Although educational inquiry does not have a generally accepted, explicitly formulated and fully comprehensive scheme for weighing the evidence, it can be achieved by following the "scientific method". Remember, what is being claimed here is roughly that the "scientific method" as viewed by logical positivists is the proper method. At best, the so-called "scientific method" is only a formalization of some aspects of scientific methodology. Physics itself could not proceed without unformalized, practical knowledge (Putnam, p. 72. See also Brown, Black, Cohen, Hesse, Salmon, Thagard).

B. In Defense of Imaginative Participation

Let's define a personal perspective as the system of concepts, axiomatic assumptions and beliefs, modes of expression of feelings and attitudes, strategies for problem-solving, etc., which are, at a certain time, "given" for a particular person; the cognitive, perceptual, expressive, and volitional structures in terms of which, and by reference to which, he "sizes up" his "world" (Black, 1975a, p. 265).

Although the notion of perspective has been introduced for a particular person, this perspective intersects the perspectives of groups of people, say a group of students, or a group of scientists, or even scientists as a whole. Indeed, the "language" that is a central component of a person-

perspective, no matter now self-centered, idiosyncratic or solipsistic, is necessarily a social instrument: every utterance is potentially intelligible to another person, every expression might "make sense" to a companion (Black, 1975a, p. 262; Taylor, 1977). Person-perspectives and the group-perspectives that overlap them might be called human-perspectives. In getting a detailed knowledge of such a person-perspective, we shall focus upon the features of the perspective that are sharable, as constituting part of a group-perspective (Black, 1975, p. 265).

Earlier I have written that the central questions which criterion-referenced measurement raises are:

I) What concepts, perceptions, attitudes, beliefs, cognitive skills and expectations are characteristic or essential of a person's doing such-and-such (or a person's having the ability to do something)? (This is a test of explication, articulation and clarification.)

II) Under what conditions in what situations is it reasonable to "infer" that a person is doing such-and-such (or can do something) and what is the nature of the "inference"? (This is a question methodology.)

I want to claim that a basic method of progress toward "perceptual articulation" is imaginative participation--the ability to imagine what it would be like to be a particular person, to do a certain thing. Contrary to Nagel, Rudner and others, I want to argue that the method is of methodological as opposed to merely psychological significance in educational inquiry.

What kinds of reasons are there for believing that imaginative participation has methodological significance for education? Part of the answer lies in appreciating the point that education ought to be concerned with helping students acquire knowledge and understanding of and attitudes towards the world. In order to accomplish this task,

educators must be in position to determine whether the person has that knowledge, understanding, or attitude. Thus, the method of imaginative participation addresses the same questions which are central to criterion-referenced measurement.

Furthermore, consider the following line of reasoning by H. Putnam. After having studied Hebrew for a while, Putnam came to know that shemen means oil. I take it that human beings can acquire knowledge of such simple facts. The important question is whether this knowledge can be checked "scientifically" in Mill's or Nagel's sense? The answer--no. As Putnam has put it:

> "Testing the 'hypothesis' that shemen means oil involves testing the conjunction of an 'analytical hypotheses' I can't state properly (as the failure of mechanical translation shows) and psychological premises I obviously can't list in advance! In fine, knowledge of such a simple fact as 'shemen means oil' cannot be justified/confirmed by following the paradigms of inductive logic (or SM)." (p. 70)

Is this an unusual case? No. One's knowledge that he is talking to people and understanding their answers is typical practical knowledge. The inferences are not, in any serious non-trivial sense, logical positivistic scientific inferences. And the key terms--"means", "understand"--cannot, as Putnam has argued, be made scientifically precise (p. 70). Therefore, Rudner and others are quite wrong when they hold that a method such as "imaginative participation" must be supplemented by the"scientific method" (1966, pp. 72-73). (See also Salmon and Thagard.)

Putnam has argued that what we "obtain" by the method of imaginative participation needs to be checked and so is not knowledge. Yet it doesn't follow it is only "hypothesis". Imaginative participation may give less than knowledge (with a capital "K"), yet it gives more than mere logical or physical possibility. Imaginative participation gives

plausibility--it is the source of prior probability in many judgments about people. As I have noted above, it is widely accepted doctrine in Philosophy of Science that inductive testing of theories presupposes some a priori weighing of the theories. Hypothesis testing in science requires a Bayesian methodology. It is also accepted that no one has been able to formalize completely the intuitive judgments of scientists which in practice constitute "the prior probability metric." (But see Cohen, Hesse, Levi, Salmon and Thagard.)

Thus, the source of the prior probability metric in many judgments about people is the method of imaginative participation. The "hypothesis" which we assign a significantly higher probability on the basis of the method (IP) must be checked, but even this checking is ultimately intuitive. In Platonic terminology, it may not provide "knowledge", but it does provide "right opinion". And as Putnam has argued, knowledge depends on a good deal of "right opinion" (pp. 74-75).

At this point it is important to recall the differences between mathematical probability and inductive probability. (See III.D.) Although Putnam has argued that a Bayesian approach is defensible (and hence, so is the application of mathematical probability), an alternative account of such reasoning and inference can be given which uses inductive probability. Given an approach where inductive probability is employed, one can reach conclusion similar to those of Putnam. (See Cohen, especially sections 68-70.) One suspects that Black's position would most accord with the approach which employs inductive probability. (1970b, 1975a.)

How is it that we have this knowledge--knowledge that cannot be "scientized"--verified by a method that publicly conforms to the canons of logical positivistic scientific methodology, knowledge that depends

on unformalized practical knowledge (what Polanyi calls "implicit knowledge")? One source is that we can acquire skills that are too complex to describe by a theory. For example, one can learn to translate from one language to another, but one cannot describe that acquired skill by an explicit theory. The fact is that skills don't always depend on theories (Consider the example of walking). Knowledge--even verbalized knowledge--can be embodied in a skill and not a theory (Putnam, p. 71).

Thus, one is lead to the following striking point: <u>What is teachable need not be explicitly formalizable</u>.

It is important to note that ordinary, everyday psychological explanations of people's actions are another example of the same phenomena. Although it may be perfectly clear to everyone in a given situation that Jones is jealous of Smith's reputation, one couldn't give anything like a "scientific proof" of it (Putnam, pp. 71-72).

It is natural, especially for scholars who are professional writers and talkers, to think immediately of verbalizing, of formalizing the perspectival apparatus. But, as Max Black has argued, throughout ordinary life there is <u>radical untranslatability</u>:

> "The difficulty is partly in the 'higher multiplicity' of the action vis a' vis its verbal description: the near impossibility of capturing in words the difference between the way a child strokes a cat and its mother does. But it is not only that. So far as it goes, the expression 'a jealous retort' may be perfectly correct in its application to a particular act and yet, as we say, mean nothing, to somebody who has never been jealous, for whom jealous is only a name. (There is such a thing as feeling blindness as well as color blindness.) To the extent that actions are the enactment of attitudes, we can fully understand them only by some process of imaginative participation...in which, by rehearsing the attitude, we come to know it from the inside." (p. 266)

Does it follow that because the rhythmic and qualitative aspects of action are not to be captured in verbal description that they are

inexpressible and somehow ineffable? No. They are typically expressed, and the attitudes, motives, intentions, etc., that inform them are essentially expressible. (Black, p. 266)

What should we call what we obtain by imaginative participation? In contrast to Putnam, Black has argued that knowledge is a plausible answer--the kind of knowledge that we have when we understand, perhaps after reading some fictional depiction of jealous actions, what jealousy is like. Those who would deny this the name of knowledge are in the grip of an ancient prejudice that insists on restricting knowledge to what can be talked about (p. 267). (See also Cohen.)

If we use the term "humanistic studies" to refer to second-order activities of "placing", interpreting, criticizing, and evaluating primary "humanistic" texts, works of art, and related artifacts, institutions, and practices, then a distinctive task of humanistic criticism is the delineation and articulation of "human perspectives" (Black, p. 262). According to Black, it would be gratuitous reductionist folly to deny the central place of imaginative participation in the enterprise of humanistic criticism (p. 266).

Putnam has argued that the logical positivistic picture of social science--as based on the "general laws" of history, sociology, and psychology", confirmed by the "SM" and applied to the data as in physics--is false to present social science and utopian as a vision of future social science (p. 75). How does a historian--or a social scientist--decide that "Smith was hungry for Power"? (or that Jones was jealous of Smith's reputation?) Not by applying the "general laws" of history, sociology, and psychology to the data as logical positivist methodologist urge he should!

> "Rather, he has to absorb all this material, and then rely on his human wisdom that this shows power-hunger 'beyond reasonable doubt' (as the courts say). In effect, he uses himself as a 'measuring instrument'; which is pretty much what Weber, Dilthey, etc., urged." (p. 73)

Putnam went on to argue that it is a feature of scientific knowledge, e.g., physics, that we use measuring instruments that we understand. Our theory applies to our measurements, and to their interactions with what they are used to measure, not just to the objects we measure. It is important to notice, however, that

> "it is a feature of practical knowledge that we often have to use ourselves (or other people) as the measuring instruments--and we do not have an explicit theory of these interactions." (p. 72)

If we want to determine whether a student knows a certain thing about physics, or whether a student understands, or whether a student has a certain attitude towards physics, we'll need to have educators -teachers - who can make these determinations, people who can act as our measuring instruments.

This is an important point: Our best measuring instruments may well be (knowledgeable) human beings.

As I have argued earlier, criterion-referenced measurement takes as a basic both the task of articulation and clarification and and the task of methodology. Here I have tried to show that imaginative participation ought to be a basic method in criterion-referenced measurement. I do not deny that many questions remain to be answered. Of great importance are the questions where and why is a Bayesian account which employs mathematical probability justifiable and where and why is an account which employs inductive probability justifiable. It is time, however, that educators gave up their absolute reliance on the logical positivistic

"scientific method". The method of imaginative participation gives an alternative for which there are very good reasons to believe that it can generate knowledge and understanding of educational activities.

V. Concluding Remarks

It has been my belief throughout this essay that Criterion-Referenced Measurement theory sets out to answer questions and to perform tasks which are profoundly important for education. Roughly, the theory sets out to determine whether a person can (or can't) do a certain thing and to establish and justify the method (or methods) by which such determinations can be made. Such questions and tasks essentially involve a matter of articulation and clarification and a matter of methodology.

In this essay I have addressed these basic aspects of current CRM theory. I have had reason to recommend that certain concepts and methods be rejected while certain other concepts and methods be employed. I shall not repeat a summary of the points and issues here (see Section I).

I would like to add, however, that recent efforts in CRM theory are quite provoking and stimulating. The questions raised and the methodologies recommended require the attention of some of our best scholars and practitioners. And I believe this task must be an interdisciplinary one--philosophers, psychologists, sociologists, psychometricians and those who-have-the-ability will need to be involved. For example, we have yet to understand fully how it is that scientists determine their prior probability metric, not yet do we understand fully how probability attributions are employed in everyday, practical-ability statements.

FOOTNOTES

[1]In what follows I shall concentrate on the concept of ability. Of course, analyses of the concept of attitude and the concept of belief will lead to results different from the analysis built on the concept of ability.

[2]For example, I. Levi adopts a generalized version of a maximin criterion (p. 145).

BIBLIOGRAPHY

Achinstein, P. Concepts of Science. Baltimore: Johns Hopkins Press, 1969.

Aune, B. Can. In P. Edwards, (ed.) The encyclopedia of philosophy. New York: Macmillan, 1967.

Aune, B. Knowledge, mind, and nature. Reseda, CA, Ridgeview Publishing Co., 1967.

Berofsky, B. "Book Review of The Cement of the Universe: A Study of Causation." Journal of Philosophy, 74 (1977), pp. 103-118.

Black, M. Induction. In Margins of precision: Essays in logic and language. Ithaca, NY: Cornell University Press, 1970a, Chapter V, pp 57-90. (Also published in P. Edwards, (Ed.) The encyclopedia of philosophy. New York: Macmillan, 1967.

Black, M. Is scientific neutrality a myth. Public lecture at the Annual Meeting of the American Association for the Advancement of Science, January 27, 1975B, New York.

Black, M. Probability. In Margins of precision, 1970b, Chapter VI, pp. 91-136. (Also published in The encyclopedia of philosophy, MacMillan, 1967.)

Black, M. Some half-baked thoughts about induction. In The margins of precision, 1970c, pp. 137-144.

Black, M. Some tasks for the humanities. In Caveats and critiques: philosophical essays in language, logic and art. Ithaca, NY: Cornell University Press, 1975a.

Black, N. "Psychologism and Behaviorism", Philosophical Review, C, No.1, Jan 1981.

Brown, H.I. Perception, Theory, and Commitment, University of Chicago Press, 1977.

Chisholm, R. Perceiving Ithaca, N.Y.: Cornell University Press, 1957.

Cohen, L.J. The Probable and the Provable. Oxford: Clarendon Press, 1977.

Cronbach, L. J. Test validation. In R. L. Thorndike (Ed.), Educational measurement. (2nd ed.). Washington, D.C.: American Council on Education, 1971, pp. 443-507.

Davis, L. H. Theory of action. Englewood Cliffs, NJ: Prentice-Hall, 1979.

Ellett, F.S., Jr. "Reflections on the Content-Construct Validity Distinction in Educational Test Theory", Proceedings of the Far Western Philosophy of Education Society, December, 1980.

de Finetti, B. Probability, Induction, and Statistics. New York; Wiley, 1972.

Foster, J.A. "Testing the Cement: An Examination of Mackie on Causation," Inquiry, 18 (1975), pp. 487-498.

Geach, P. Mental Acts, (Routledge, London, 1956).

Hamlyn, D.W. "Human Learning", in R.S. Peters, (ed.) The Philosphy of Education, (Oxford University Press, 1973), 178-194.

Hesse, M. "Bayesian Methods and the Initial Probabilities of Theories", in Maxwell and Anderson, Jr. (eds.), Induction, Probability and Confirmation. Vol.VI. Minnesota Studies in the Philosophy of Science. Minnesota: University of Minnesota, Press, 1975. 50-105.

Hively, W., et al. Domain-referenced curriculum evaluation: A technical handbook and a case study from the Minnemast Project. CSE Monograph Series in Evaluation #1. Center for the Study of Evaluation, University of California, Los Angeles, 1973.

Kenny, A. Action, emotion, and will. New York: Humanities Press, 1963.

Lewis, D. "Counterfactual Dependence and Time's Arrow," Nous 13 (1979), 455-476.

Linn, R. L. Conceptualization of issues in construct and content validity. Center for the Study of Evaluation, University of California, Los Angeles, June, 1978.

Malcolm, N. Behaviorism as a philosophy of psychology. In T. W. Wann, Behaviorism and phenomenology. Chicago: University of Chicago Press, 1964, pp. 141-154.

Malcolm, N. Problems of mind. New York: Harper Torchbooks, 1971.

Mill, J. S. A system of logic. (8th ed.). New York, 1881.

Millman, J. Passing scores and test lengths for domain-referenced measures. Review of Educational Research, 1973, 43, 205-216.

Millman, J. Criterion-referenced measurement. In W. J. Popham (Ed.), Evaluation in education: Current applications. Berkeley, CA: McCutchen Pub. Co., 1974.

Nagel, E. The logic of historical analysis. In A. Feigl & M. Brodbeck (Eds.), Readings in the philosophy of science. New York: Appleton-Century-Crofts, 1953, pp. 685-700.

Nagel, E. The structure of science, New York: Harcourt, Brace & World, 1961.

Novick, M. R. & Lewis, C. Prescribing test length for criterion-referenced measruement. In C. W. Harris, M. C. Alkin and W. J. Popham (Eds.), Problems in criterion-referenced measurement, CSE Monograph Series in Evaluation #3, Center for the Study of Evaluation, University of California, Los Angeles, 1974, pp. 139-158.

Popham, W. J. Domain specification strategies. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art, Baltimore, Maryland: The John Hopkins University Press, 1980b, pp. 15-31.

Popham, W. J. & Lindheim, E. The practical side of criterion-referenced test development, NCME: Measurement in Education, Vol. 10, No. 4, Spring, 1980a, pp. 1-7.

Popper, K. Objective Knowledge (Oxford: Clarendon Press, 1972).

Putnam, H. "Logical Positivisen and Philosphy of Mind", in Achinstein and Barker (eds.) The Legacy of Logical Positivim, Baltimore: John Hopkins, 1969, pp. 211-225.

Putnam, H. Meaning and the moral sciences. London: Routledge and Kegan Paul, 1978.

Putnam, H. Degree of Confirmation and Inductive Logic. In Mathematics, Matter, and Method: Philosophical Papers, Vol.1, Cambridge University Press, 1975a, chapter 17, pp.270-292.

Putnam, H. What theories are not. In Mathematics, Matter and Method. Philosophical Papers, Vol. 1, Cambridge University Press, 1975, Chapter 13, pp. 215-227.

Richards, D. A. J. A theory of reasons for action. Oxford University Press, 1971.

Rudner, R. Philosophy of social science. Englewood Cliffs, NJ: Prentice-Hall, 1966.

Salmon, W.C. "Baye's Theorem and the History of Science", in R.H. Stuewer, (ed.) Historical and Philosophical Perspectives of Science Minnesota Studies in the Philosphy of Science, Vol.V, Minnesota: University of Minnesota Press, 1970, pp. 68-86.

Savage, L.J. The Foundations of Statistics, (New York: Wiley, 1954).

Scheffler, I. A note on behaviorism as an educational theory. In Reason and Teaching. Indianapolis: Bobbs-Merrill, 1973, pp. 165-170.

Scriven, M. A study of radical behaviorism. In Minnesota studies in the philosophy of science, Vol. 1, Minneapolis: University of Minnesota Press, 1956.

Taylor, C. Interpretation and the sciences of man. In F. R. Dallmayr & T. A. McCarthy (Eds.), Understanding and social inquiry. Notre Dame: University of Notre Dame, 1977, pp. 101-131.

Thagard, P.R. "The Best Explanation: Criteria for Theory Choice", Journal of Philosphy, Feb. 1978, 76-92.

Toulmin, S. Human Understanding, Princeton: Princeton University Press, 1971.

Van Fraassen, B.C. The Scientific Image, Oxford: Clarendon Press, 1980.